

Analisis Sentimen Aplikasi Maskapai Penerbangan Lion Air Menggunakan Metode SVM dan Naïve Bayes

Risa Sulistiawati¹, Mia Kamayani Sulaeman²

risa201595@gmail.com, mia.kamayani@uhamka.ac.id

Universitas Muhammadiyah Prof. Dr. Hamka

Informasi Artikel	Abstrak
Diterima : 30 Mar 2024 Direview : 8 Mei 2024 Disetujui : 15 Jun 2024	Aplikasi Lion Air merupakan sebuah aplikasi pembelian tiket pesawat yang diluncurkan pada tanggal 21 Oktober 2014. Dapat diunduh dan digunakan di mana saja, kapan saja. Aplikasi Lion Air ini tersedia di Google Play Store dan juga Appstore, yang tujuannya untuk memudahkan pengguna dalam melakukan proses pembelian tiket pesawat secara online. Didalam beberapa artikel berita yang melaporkan bahwa lion air merupakan maskapai penerbangan terburuk di dunia. Namun perlu disadari aplikasi lion Air juga memiliki banyak pengguna yang memberikan ulasan positif, negatif maupun netral dikarenakan adanya beberapa ulasan yang dipaparkan di aplikasi Play Store. Masalah ini diteliti untuk analisis sentimen untuk mendapatkan rating kepuasan pelanggan kepada aplikasi Lion Air dengan perolehan 2000 data. Dalam penelitian ini, perhitungan Support Vector Machine (SVM) dan perhitungan Naive Bayes dibandingkan dengan menggunakan 80% rasio latih dan 20% rasio uji. Dalam pertimbangan ini digunakan 795 opini positif dan 805 opini negatif yang digunakan, dimana Support Vector Machine (SVM) dengan fitur Bigram menjadi metode paling unggul dengan presisi 99,23%, recall 83,03%, akurasi 91,75%, Skor F-1 sebesar 90,51%.

Keywords	Abstract
Sentiment Analysis, SVM, Naïve Bayes, Lion Air Application, Google Playstore	<i>Lion Air App is a flight ticket purchase application launched on October 21, 2014. It can be downloaded and used anywhere, anytime. Lion Air App application is available on the Google Play Store and also the Appstore, which aims to facilitate users in the process of purchasing airplane tickets online. In several news articles reporting that Lion Air is the world's worst airline. in the world. However, it needs to be realized that the Lion Air application also has many users who give positive, negative and neutral reviews due to several factors. neutral due to the existence of several reviews presented in the Play Store application. This problem was researched for sentiment analysis to get a customer satisfaction rating for the Lion Air application. Lion Air application with the acquisition of 2000 data. In this research, Support Vector Machine (SVM) calculation and Naive Bayes calculation were compared using 80% training ratio and 20% test ratio. In this consideration, 795 positive opinions and 805 negative opinions were used. used, where Support Vector Machine (SVM) with Bigram features became the most superior method with 99.23% precision. method with 99.23% precision, 83.03% recall, 91.75% accuracy, F-1 score of 90.51%.</i>

A. Pendahuluan

Didunia saat ini, perkembangan teknologi sangatlah pesat[1], yang membuat kehidupan manusia sangat bergantung terhadap penggunaan teknologi, hampir semua bidang saat ini menggunakan teknologi termasuk dalam pembelian tiket penerbangan melalui aplikasi, saat ini pembelian tiket secara online menjadi lebih diminati oleh masyarakat, kemudahan dan efektifitas menjadi pengaruh adanya sistem pembelian tiket secara online melalui aplikasi. Hal tersebut mendorong beberapa maskapai mengikuti perkembangan teknologi dengan cara menyediakan Fly Booking App.

Fly Booking App adalah aplikasi yang digunakan untuk melakukan pembelian tiket penerbangan secara online agar memudahkan pengguna dalam proses transaksi. Sesuai yang akan terjadi, Google dan Temasek *Research e-Conomy SEA 2019* Ekonomi Internet di Asia Tenggara Mencapai Titik Balik Pasar layanan tiket online GBV di Asia Tenggara telah mencapai \$29,7 miliar. Nilai ini merupakan yang terbesar dibandingkan sektor ekonomi Internet lainnya, seperti *e-commerce* (\$23 miliar), media online (\$11 miliar), dan lalu lintas online (\$8 miliar)[2].

Traveloka adalah aplikasi *Fly Booking App* merupakan aplikasi yang paling sering digunakan para pengguna ponsel. Hal itu diakibatkan aplikasi Traveloka mempunyai interface sederhana sehingga mampu digunakan oleh orang awam sekali pun [3]. Salah satu maskapai penyedia *Fly Booking App* adalah *Lion Air*, Aplikasi *Lion Air* di luncurkan pada tanggal 21 Oktober 2014 berguna untuk memudahkan pengguna dalam melakukan proses pembelian tiket pesawat secara online. Didalam beberapa artikel berita yang melaporkan bahwa lion air merupakan maskapai penerbangan terburuk di dunia dalam survey perjalanan yang dilakukan oleh Travel Bounce, pada sebuah platform Australia [4]. Namun perlu disadari aplikasi lion Air juga memiliki banyak pengguna yang memberikan ulasan positif, negatif serta netral pada salah satu toko aplikasi terbesar yang menyediakan jutaan aplikasi yaitu *Google Play Store*.

Google Play adalah manfaat penyebaran terkomputerisasi yang didukung dan dibuat oleh Google. Ini adalah toko aplikasi resmi untuk sistem operasi Android yang memungkinkan pengguna menelusuri dan mengunduh aplikasi yang dibuat menggunakan Unit Peningkatan Program Android (SDK) dan dirilis oleh Google[5]. Data review aplikasi *Lion Air* diperoleh dengan menggunakan teknik Web Scraping dengan bahasa pemrograman Python. Teknik penangkapan online digunakan ketika memasukkan beberapa kode, dan data dikumpulkan sebagai file CSV, yang dapat digunakan sebagai sumber informasi bagi peneliti ketika melakukan analisis sentimen. [6].

Penelitian ini berfokus pada analisis sentimen melalui review pengguna aplikasi *Lion Air* terkait *booking* tiket pesawat, guna menunjukkan betapa pentingnya sebuah kualitas layanan walaupun dalam sistem pembelian secara online, hal ini dapat berdampak pada tingkat kepercayaan pengguna dan juga kebutuhan pengguna aplikasi *Lion Air*.

Analisis sentimen adalah penggunaan analisis konten tekstual untuk mengekstraksi berbagai sumber fakta dari internet dan berbagai sistem media sosial. Tujuannya adalah untuk mendapatkan kritik dari pengguna platform [7]. Analisis sentiment dapat dilakukan dengan beberapa metode yaitu, Decision Trees,

Support Vector Machine, k-Nearest Neighbor (k-NN), Naïve Bayes, Random Forest, Logistic Regression dan Linear Regression.

Penelitian analisis sentimen sebelumnya pernah dilakukan oleh Gracia Radiena dan Adi Nugroho, penelitiannya menggunakan data komentar aplikasi KAI Access di Google Play Store. Metode yang digunakan adalah Support Vector Machine (SVM). menggunakan SMOTE. Pada preprocessing dilakukan Transform Case, Tokenizing, Stemming dan Stopwords. Hasil penelitian yang dari beberapa aspek yang didapatkan adalah aspek Efficiency sebesar 94,38%, aspek Learnability sebesar 94,73%, aspek Satisfaction sebesar 87,26% dan aspek Errors sebesar 85,13%, [8].

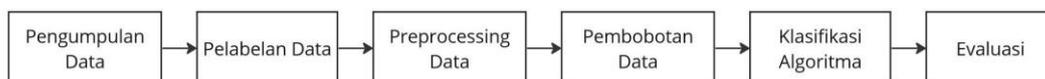
Penelitian lain menggunakan algoritma Support Vector Machine dilakukan oleh Alman Muhammadin dkk. Alman Muhammadin dkk menggunakan 1000 data ulasan pada aplikasi Kredivo menggunakan dua perhitungan, yaitu Support Vector Machine (SVM) dan Naive Bayes Clacifier (NBC). Perhitungan Naïve Bayes Classifier memperoleh nilai ketepatan sebesar 80,8%, sedangkan perhitungan Back Vector Machine (SVM) memperoleh nilai ketepatan sebesar 83,3%. [9].

Penelitian analisis sentimen juga pernah dilakukan oleh Tia Anggita Sari dkk, Tia dkk melakukan penelitian analisis sentiment terhadap aplikasi Fore Coffe menggunakan metode Naïve Bayes, didalam penelitiannya Tia dkk menggunakan 1801 data ulasan dan opini yang terdaftar mayoritas positif sebanyak 1163, sedangkan opini negatif hanya 315. Nilai akurasi yang diperoleh dari penelitian ini sebesar 74,28%, recall 45.33%, dan precision 24.46% [10].

Berdasarkan uraian diatas, penulis akan meneliti analisis sentiment ulasan pada aplikasi Lion Air dengan menggunakan metode Support Vector Machine (SVM) dan Naïve Bayes dimana kelas sentiment dibagi kedalam ulasan positif dan negatif. Evaluasi yang digunakan yaitu akurasi, recall, perisi, dan f-measure. Data set diperoleh melalui teknik web scrapping dan pada teks dilakukan preprocessing sebelum pengklasifikasi. Tujuan penelitian ini diharapkan menjadi evaluasi bagi PT. Lion Group untuk meningkatkan kualitas aplikasi mobile Lion Air serta diharapkan dari penelitian ini dapat mengetahui hasil analisis sentiment pada kelas positif, dan negatif.

B. Metode Penelitian

Pada penelitian ini metode penelitian yang digunakan termasuk proses untuk menganalisis sentimen pengguna aplikasi Lion Air terhadap kepuasan pengguna aplikasi melalui ulasan Google Playstore. Metode Penelitian yang digunakan peneliti terbagi menjadi beberapa tahap yaitu:



Gambar 1. Alur Penelitian

1. Pengumpulan Data



Gambar 2. Proses penarikan dataset

Tahap awal yang dilakukan penulis adalah melakukan pengumpulan data atau crawling data, metode yang penulis gunakan untuk pengumpulan data adalah metode web scrapping, web scrapping adalah program yang dibangun untuk mengumpulkan data langsung dari kode sumber situs web, baik itu HTML, CSS, atau informasi lainnya di situs web [11]. tools yang digunakan penulis untuk melakukan web scrapping adalah google colab dengan menggunakan bahasa pemrograman python.

2. Pelabelan Data

Penyusunan labeling pada dataset ini dilakukan secara fisik oleh tiga orang mahasiswa Program Pendidikan Bahasa dan Sastra Indonesia di Perguruan Tinggi Muhammadiyah. Prof Dr Hamka mengevaluasi survei klien dalam dua kelas opini, yaitu positif dan negatif menggunakan teknik voting.

3. Preprocessing Data

Suatu kata tidak dapat langsung diproses oleh algoritma pencarian, sehingga harus dilakukan preprocessing teks untuk mengubah teks menjadi data numerik. Pada fase ini, data yang telah disiapkan diubah menjadi data yang siap untuk dianalisis [12].

4. Pembobotan Data

Untuk melakukan pembobotan data atau ekstraksi fitur pada penelitian yang dilakukan ini, penulis menggunakan metode TF-IDF dan Bigram. Dimana TF-IDF merupakan strategi yang dapat menghitung bobot setiap kata dan sering digunakan dalam pemulihan data pemulihan. Fitur ini juga dikenal mahir, sangat mendasar, sederhana dan memberikan hasil yang tepat [13] Pada algoritma TF-IDF digunakan rumus untuk menghitung bobot (W) masing masing dokumen terhadap kata kunci dengan rumus yaitu :

$$W_{dt} = tf_{dt} * Idf_t \quad (1)$$

$$IDF = \log \left(\frac{N}{df} \right) \quad (2)$$

Dimana:

- d = dokumen ke-d
- t = kata ke-t dari kata kunci
- W = bobot dokumen ke-d terhadap kata ke-t
- tf = banyaknya kata yang dicari pada sebuah dokumen

- IDF = Inversed Document Frequency
- N = total dokumen
- df = banyak dokumen yang mengandung kata yang dicari

. Sementara itu, Bigram adalah N-gram berukuran dua di mana N-gram dapat merupakan proses yang digunakan untuk penambangan teks dan pemrosesan bahasa, yang menghasilkan hingga n-karakter dari sebuah string. Pada pembagian kata, strategi ini dilakukan dengan mengisolasi sebanyak n kata dari susunan kata (bagian, kalimat, masukan) yang diteliti secara persisten dari awal hingga akhir laporan [14].

Tabel 1. Ilustrasi Pembobotan Bigram

Sebelum Pembobotan Bigram	Setelah Pembobotan Bigram
aplikasi tidak jelas tidak ada niat developer untuk meningkatkan kinerja aplikasi	“aplikasi tidak”, “tidak jelas”, “jelas tidak”, “tidak ada”, “ada niat”, “niat developer”, “developer untuk”, “untuk meningkatkan”, “meningkatkan aplikasi”
pindah bahasa susah tikecting susah	“pindah bahasa”, “bahasa susah”, susah ticketing”, “ticketing susah”

5. Klasifikasi Algoritma

Pada tahap ini peneliti menggunakan Pada tahap implementasi algoritma, penulis menerapkan dua algoritma untuk model klasifikasi yaitu algoritma *Support Vector Machine* (SVM) dan algoritma Naïve Bayes.

5.1 Algoritma Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma klasifikasi dengan metode yang masih cukup baru yang digunakan untuk menciptakan ekspektasi dalam klasifikasi. Perhitungan ini termasuk dalam kategori Supervised learning, karena dalam pelaksanaannya SVM memerlukan sesi persiapan dengan menggunakan persiapan SVM berturut-turut yang dilanjutkan dengan sesi pengujian.[9].

$$f(x) = \text{sign}(w \cdot x + b) \quad (3)$$

di mana:

- $F(X)$ = adalah fungsi keputusan (output klasifikasi),
- w = adalah vektor bobot,
- x = adalah vektor fitur input,
- b = adalah bias,
- \cdot = menunjukkan operasi dot product.

5.2 Algoritma Naïve Bayes

Bayes merupakan metode prediksi berbasis probabilitas yang menggunakan prinsip Bayes dan aturan Bayes berdasarkan asumsi independensi yang ketat (non-dependensi). Dengan kata lain, Naïve Bayes adalah model khusus atau model fitur independen[15]. Prediksi Estimasi Bayes didasarkan pada teori Bayes yang menggunakan model persamaan umum[16] :

$$P(Y|X) = \frac{P(X|Y)P(H)}{P(E)} \quad (4)$$

Dimana :

- $P(Y|X)$ = Probabilitas bebas bersyarat (conditional probability)
- $P(X|Y)$ = Probabilitas sebuah bukti
- $P(Y)$ = Probabilitas kelas awal Y
- $P(X)$ = Probabilitas awal (priori)

6. Evaluasi

Evaluasi dilakukan dengan menggunakan metrix evaluasi seperti recall, presisi, akurasi, dan F-measure. Metode evaluasi ini membantu untuk menilai seberapa baik model dalam mengklasifikasikan sentimen teks, apakah positif, negatif, atau netral. Untuk melakukan evaluasi model dapat dilakukan menggunakan teknik *confusion matrix*. Dibawah ini merupakan model dari table *confusion matrix* [17].

Tabel 2. Confusion Matrix

Actual Class	Prediction Class	
	positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Dari tabel di atas, nilai dari recall, presisi, akurasi dan F1 score dapat diperoleh. Akurasi digunakan untuk menilai kinerja model dalam memprediksi sentimen dari teks. Model yang memiliki akurasi yang tinggi menunjukkan bahwa mereka efektif dalam memprediksi sentimen dari data yang tidak diketahui [18]. recall mengukur seberapa efisien model tersebut dalam mengidentifikasi keseluruhan sentimen positif, negatif, atau netral dari teks. Precision atau presisi mengukur seberapa baik model dalam mengidentifikasi teks yang sebenarnya positif dari keseluruhan teks yang diprediksi sebagai positif. Sedangkan F-measure merupakan rata-rata harmonis dari presisi dan recall model.

$$accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (5)$$

$$recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (6)$$

$$\text{precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (7)$$

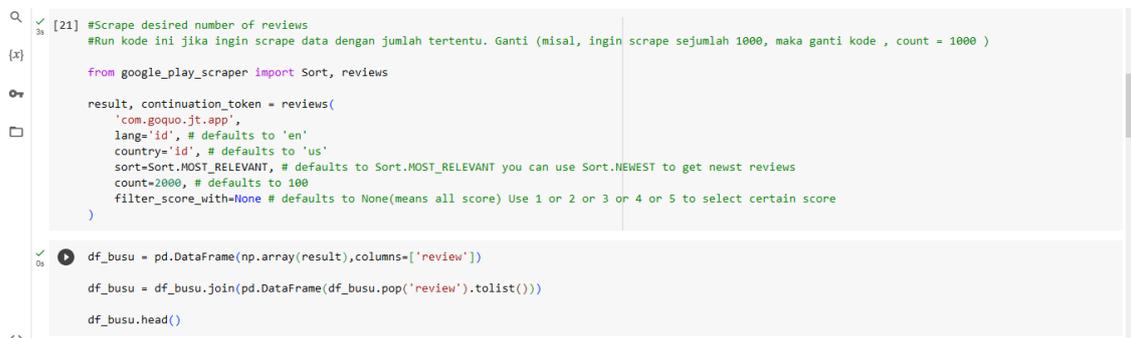
$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (8)$$

C. Hasil dan Pembahasan

Segmen ini berisi pertanyaan tentang hasil yang diperoleh dengan menggunakan perhitungan Support Vector Machine (SVM) dan Naive Bayes.

1. Pengumpulan Data

Pada tahap ini, peneliti menggunakan teknik web scrapping untuk proses pengambilan data melalui Playstore. Hasil yang diperoleh oleh peneliti dari proses web scrapping ini dihasilkan 2000 data ulasan dimana informasi yang diambil merupakan informasi terpenting dan terkini.



```
[21] #Scrape desired number of reviews
#Run kode ini jika ingin scrape data dengan jumlah tertentu. Ganti (misal, ingin scrape sejumlah 1000, maka ganti kode , count = 1000 )

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.goquo.jt.app',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT you can use Sort.NEWEST to get newest reviews
    count=2000, # defaults to 100
    filter_score_with=None # defaults to None (means all score) Use 1 or 2 or 3 or 4 or 5 to select certain score
)

df_busu = pd.DataFrame(np.array(result), columns=['review'])

df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))

df_busu.head()
```

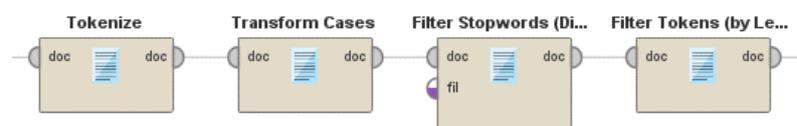
Gambar 3.Scraping Data

2. Pelabelan Data

Langkah selanjutnya yang dilakukan peneliti adalah memberi label pada data. Peneliti dengan mengumpulkan 1.600 data yang dilabel oleh tiga mahasiswa program Studi Pendidikan Bahasa dan Sastra Indonesia Universitas Muhammadiyah Prof. Dr. Hamka dan menghasilkan 805 data negatif dan 795 data positif dengan menggunakan teknik voting.

3. Preprocessing Data

Pada preprocessing data dilakukan tahap proses pembersihan data serta process documents from data yang didalamnya berisi proses *tokenize*, *transform cases*, *filter stopwords* dan *filter tokens by length*.



Gambar 4.Proses Preprocessing Data

Pada *tokenize*, teks dibagi menjadi bagian yang lebih kecil, seperti frasa atau kata-kata. Untuk mempermudah analisis struktur kalimat.

Tabel 3. Ilustrasi *Tokenize*

<i>Sebelum Tokenize</i>	<i>Sesudah Tokenize</i>
Perubahan penerbangan tidak ada konfirmasi ke customer Cs ditanya masalah kompensasi apa malah bingung	'Perubahan', 'penerbangan', 'tidak', 'ada', 'konfirmasi', 'ke', 'customer', 'Cs', 'ditanya', 'masalah', 'kompensasi', 'apa', 'malah', 'bingung'
Tapi ini sangat bagus untuk beli tiket pesawat kecuali bisa	'Tapi', 'ini', 'sangat', 'bagus', 'untuk', 'beli', 'tiket', 'pesawat', 'kecuali', 'bisa'

Pada tahap *transform cases*, kata pada data set yang berawalan huruf kapital diubah menjadi huruf kecil.

Tabel 4. Ilustrasi *Transform Cases*

<i>Sebelum Transform Cases</i>	<i>Setelah Transform Cases</i>
BAD SERVICE CANCEL SCHEDULE SEENAK JIDATDAN NOTIF VIA EMAIL PUN GA JELAS GA ADA INFO YG LENGKAP CS PUN SUSAH DIHUBUNGI GA LAGI-LAGI NAIK MASKAPAI INI KAPOK	bad service cancel schedule seenak jidat dan notif via email pun ga jelas ga ada info yg lengkap cs pun susah dihubungi ga lagilagi naik maskapai ini kapok
BAGAIMANA CARA MERUBAH NAMA TIKET YANG SUDAH SAYA PESAN	bagaimana cara merubah nama tiket yang sudah saya pesan

Pada filter stopword, akan dihilangkan kata-kata umum yang tidak memberikan banyak informasi, seperti "dan," "atau," dan "yang."

Tabel 5. Ilustrasi Penghapusan *Stopwords*

<i>Sebelum Penghapusan Stopwords</i>	<i>Setelah Penghapusan Stopwords</i>
Adanya perubahan semakin menurun Biasanya cari soekarnohatta Tinggal ketik soe aja langsung muncul Ini harus scrool sampai kebawah	perubahan semakin menurun cari soekarnohatta Tinggal ketik soe langsung muncul scrool kebawah
Aduh gimana siii Setiap saya masuk aplikasi kok keluar sendiri, padahal saya mau liat harga, tolong dong tingkatkann.	Aduh gimana siii saya masuk aplikasi keluar padahal saya liat harga tolong tingkatkann

Pada proses *filter token by length*, data yang memiliki kata kurang dari 4 karakter dan lebih dari 25 karakter akan dihapus.

Tabel 6. Ilustrasi *Filter Token by Length*

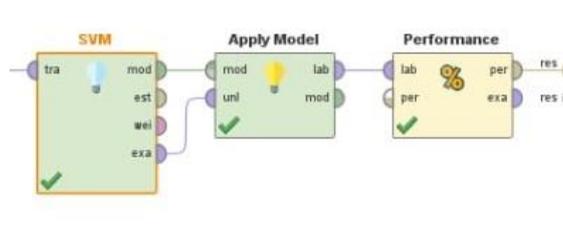
Sebelum <i>Filter Token by Length</i>	Sesudah <i>Filter Token by Length</i>
masih ada kendala untuk chekin	masih kendala untuk chekin
Mantap cek in bisa lewat hp	Mantap bisa lewat

4. Pembobotan Fitur

Pada tahap ini kata-kata pada ulasan aplikasi Lion Air diberi bobot menggunakan Fitur TF-IDF dan Bigram yang kemudian dilakukan proses preprocessing dengan tujuan untuk membuat generasi vektor menggunakan operator dokumen proses yang dibuat sebelumnya.

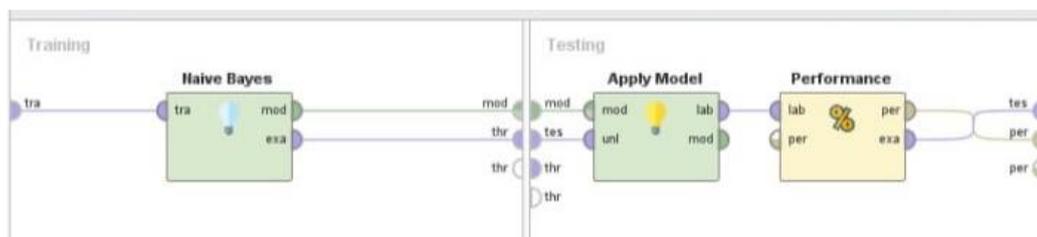
5. Klasifikasi Algoritma

Dalam proses ini, untuk melakukan analisis sentimen digunakan dua buah algoritma penghitungan yang kredibel yaitu algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan aplikasi Rapid Miner. Pada gambar 5 data berlabel dihubungkan dengan algoritma Support Vector Machine untuk mendapatkan hasil evaluasi menggunakan operator Performance dan operator Apply Model.



Gambar 5. Model Klasifikasi SVM

Selain itu, pada algoritma Naïve Bayes yang dapat dilihat pada gambar 7 digunakan operator cross validation untuk melakukan eksekusi algoritma naïve bayes, operator Apply Model dan operator Performance untuk memperoleh hasil evaluasi



Gambar 6. cross validation

6. Evaluasi Model

Setelah semua administrator yang digunakan selesai melakukan eksekusi, maka dihasilkanlah hasil dari kedua perhitungan tersebut yaitu perhitungan Naive Bayes dan SVM melalui pegangan klasifikasi menggunakan rapid miner dapat dilihat pada

gambar dibawah ini, dimana perhitungan SVM menggunakan fitur TF-IDF tersebut mendapatkan Akurasi sebesar 88,34%, recall 75,24%, presisi 99,86% dan skor f1 85,85%.

	true Positif	true Negatif
pred. Positif	1061	232
pred. Negatif	1	705

Gambar 7.Confusion Matrix SVM dengan fitur TF-IDF

Selain itu hasil evaluasi dari algoritma SVM yang menggunakan fitur bigram dapat dilihat pada gambar 8, dimana akurasi yang dihasilkan memperoleh skor 91.75%, presisi memperoleh skor 99.23%, recall 83.03% dan f-measure memperoleh skor 90.51%.

	true Positif	true Negatif
pred. Positif	1056	159
pred. Negatif	6	778

Gambar 8. Confusion Matrix SVM dengan fitur Bigram

Sedangkan perhitungan Naïve Bayes menggunakan fitur pembobotan TF-IDF dapat dilihat pada gambar 9, dimana pada akurasi memperoleh nilai sebesar 81,89%, recall sebesar 88,67%, precision sebesar 79,78% dan skor f-1 sebesar 83,92%.

	true Negatif	true Positif
pred. Negatif	690	121
pred. Positif	241	947

Gambar 9.Confusion Matrix Naïve Bayes dengan fitur TF-IDF

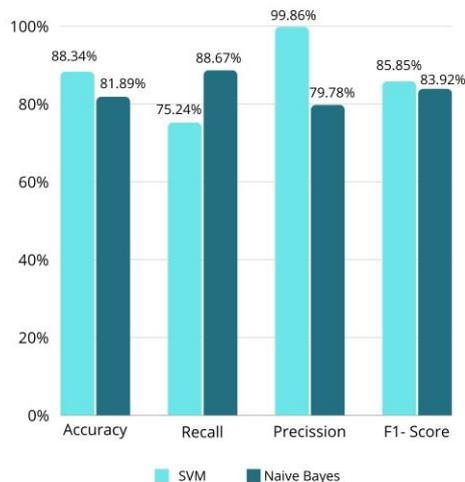
Dan juga perhitungan pada algoritma Naïve Bayes dengan fitur pembobotan Bigram dapat dilihat pada gambar 10, dimana memperoleh hasil akurasi sebesar 79.44%, presisi sebesar 80.97% recall sebesar 80.52% dan f-measure memperoleh nilai sebesar 80.71%.

	true Negatif	true Positif
pred. Negatif	728	208
pred. Positif	203	860

Gambar 10.Confusion Matrix Naïve Bayes dengan fitur Bigram

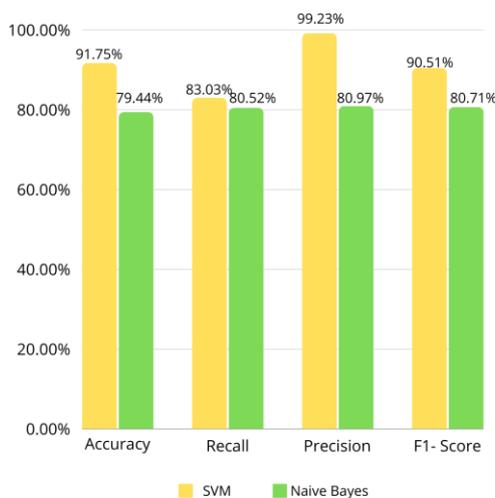
Berdasarkan hasil pengujian dari kedua algoritma dengan menggunakan fitur pembobotan TF-IDF diperoleh perbandingan yang dapat dilihat pada gambar 11,

dimana nilai Accuracy, precision dan f-1 score pada algoritma SVM lebih besar dengan memperoleh nilai sebesar 88.34%, 99.86% dan 85.85%. Sedangkan pada nilai recall algoritma Naïve Baves. memperoleh nilai yang lebih besar yaitu 88.67%.



Gambar 11. Diagram Perbandingan Nilai Evaluasi menggunakan fitur TF-IDF

Sedangkan hasil pengujian dari kedua algoritma Naïve Bayes dan SVM yang menggunakan fitur pembobotan Bigram di peroleh perbandingan yang dapat dilihat pada gambar 12, dimana nilai akurasi pada algoritma SVM masih lebih besar dari pada akurasi algoritma Naïve Bayes dimana masing-masing memperoleh 91.75% dan 79.44%, selain itu pada nilai recall, presisi dan f1- score algoritma SVM memiliki nilai yang semuanya lebih tinggi dibanding algoritma Naïve Bayes, dimana pada recall svm memperoleh nilai 83.03% sedangkan algoritma Naïve Bayes hanya memperoleh nilai 80.52%, hasil dari precision memiliki selisih yang paling jauh dimana pada algoritma SVM memperoleh nilai sebesar 99.23% dan Naïve Bayes memperoleh nilai 80.97%, dan nilai F1-Score pada algoritma SVM memiliki nilai yang lebih besar yaitu 90.51% sedangkan naïve bayes memperoleh score 80.71%.



Gambar 12. Diagram Perbandingan Nilai Evaluasi menggunakan fitur Bigram

Hasil evaluasi menggunakan algoritma Support Vector Machine dan Naïve Bayes dengan ekstraksi fitur TF-IDF dan bigram dapat diperoleh perbandingan hasil akurasi, presisi, recall, dan f1-score berbeda yang dapat dilihat pada tabel 7.

Tabel 7. Perbandingan Nilai Evaluasi

	Akurasi	Presisi	Recall	F1- Score
Naïve Bayes (TF-IDF)	81.89%	79.78%	88.67%	83.92%
SVM (TF-IDF)	88.34%	99.86%	75.24%	85.85%
Naïve Bayes (Bigram)	79.44%	80.97%	80.52%	80.71%
SVM (Bigram)	91.75%	99.23%	83.03%	90.51%

D. Simpulan

Berdasarkan hasil analisis sentimen pada maskapai penerbangan Lion Air. Dapat di ambil kesimpulan, yaitu sebanyak 2000 data komentar di goggle play store yang bertujuan untuk mengetahui tingkat kepuasan pelanggan aplikasi Lion Air, didapat 795 asumsi positif dan 805 asumsi negatif. Melihat informasi ini, nampaknya banyak orang yang kecewa dengan manfaat aplikasi tersebut. Dengan membandingkan hasil penilaian perhitungan Support Vector Machine (SVM) dan perhitungan Naive Bayes menggunakan fitur TF-IDF dan Bigram diperoleh hasil bahwa perhitungan Support Vector Machine (SVM) terbaik untuk melakukan analisis sentimen pada aplikasi pesawat Lion Air dimana pada hasil evaluasi menggunakan algoritma SVM dengan TF-IDF mendapatkan nilai presisi 88,34%, recall 75,25%, akurasi 99,86% dan f1-score 85,85%. Dibandingkan hasil dari algoritma Naïve Bayes memperoleh hasil akurasi sebesar 81,89%, presisi 79,78%, recall 88,67%, dan f1-score 83,92%. Sementara perbandingan algoritma SVM dan Naïve Bayes menggunakan fitur Bigram memperoleh hasil algoritma SVM juga lebih unggul dimana masing-masing mendapatkan akurasi 91,75%, recall 83,03%, presisi 99,23%, F1-Score 90,51%, pada algoritma SVM. Dan akurasi 79,44%, recall 80,52%, presisi 80,97%, F1-score 80,71% pada algoritma Naïve Bayes. Dari kedua algoritma tersebut dengan menggunakan dua fitur pembobotan kata TF-IDF dan Bigram, Hasil Evaluasi yang diperoleh algoritma SVM dengan menggunakan fitur pembobotan Bigram menjadi yang terbaik diantara algoritma SVM dengan fitur TF-IDF dan algoritma Naïve Bayes dengan fitur TF-IDF maupun Bigram, dimana uji evaluasi algoritma SVM dengan fitur bigram unggul pada aspek Akurasi dan F1-Score, sedangkan algoritma Naïve Bayes dengan fitur TF-IDF unggul pada aspek Recall, dan algoritma SVM dengan fitur TF-IDF unggul pada aspek Presisi.

E. Referensi

- [1] D. Salsabila, "Membongkar Rahasia Perkembangan Teknologi Yang Membuat Dunia Semakin Canggih," *FTMM NEWS*, 2023. <https://ftmm.unair.ac.id/membongkar-rahasia-perkembangan-teknologi-yang-membuat-dunia-semakin-canggih/>
- [2] Databoks, "Transaksi Tiket Online di Indonesia Diperkirakan Mencapai Rp 125 Triliun pada 2019," *Kata Media Network*, 2018. <https://databoks.katadata.co.id/datapublish/2018/12/12/transaksi-tiket-online-di-indonesia-diperkirakan-mencapai-rp-125-triliun-pada-2019>

- [3] Listiorini, "15 Aplikasi Booking Hotel dan Pesawat Terbaik di Android," *Carisinyal*, 2023. <https://carisinyal.com/aplikasi-booking-hotel-dan-pesawat/>
- [4] Atmadjati and Arista, "Maskapai Asal Indonesia Jadi yang Terburuk," *CNBC Indonesia*, 2022. <https://www.cnbcindonesia.com/opini/20221118142202-14-389312/maskapai-asal-indonesia-jadi-yang-terburuk-serius-nih>
- [5] B. Wahyudi, "Prediksi Peringkat Aplikasi di Google Play Menggunakan Metode Random Forest," *J. Nas. Teknol. Komput.*, vol. 2, no. 1, pp. 38–47, 2022, doi: 10.61306/jnastek.v2i1.25.
- [6] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," ... *Teknol. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] LP2M, "Analisis SeAnalisis Sentimen (Sentiment Analysis) : Definisi, Tipe dan Cara Kerjanya," *LP2M*, 2022. <https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/>
- [8] G. Radiana and A. Nugroho, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," *J. Pendidik. Teknol. Inf.*, vol. 6, no. 1, pp. 1–10, 2023, doi: 10.37792/jukanti.v6i1.836.
- [9] A. Muhammadin and I. A. Sobari, "Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma Svm Dan Nbc," *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 85–91, 2021, doi: 10.31294/reputasi.v2i2.785.
- [10] T. A. Sari, E. Sinduningrum, and F. Noor Hasan, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes," *Media Online*, vol. 3, no. 6, pp. 773–779, 2023, doi: 10.30865/klik.v3i6.884.
- [11] P. Andersson, "Developing a Python based web scraper," 2021.
- [12] I. D. Kurniati *et al.*, *Buku Ajar*. 2019.
- [13] Students, "About the Tutorial Copyright & Disclaimer," p. 2, 2019.
- [14] R. Ramadhan, Y. A. Sari, and P. P. Adikara, "Perbandingan Pembobotan Term Frequency-Inverse Document Frequency dan Term Frequency-Relevance Frequency terhadap Fitur N-Gram pada Analisis Sentimen," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5075–5079, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] M. A. Muslim *et al.*, *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer*, Cetakan Pe. Semarang, 2019.
- [16] F. Harahap, N. E. Saragih, E. T. Siregar, and H. Sariangisah, "Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat," *J. Ilm. Inform.*, vol. 9, no. 01, pp. 19–23, 2021, doi: 10.33884/jif.v9i01.3702.
- [17] L. Qadrini, A. Sepperwali, and A. Aina, "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1966, 2021.
- [18] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.